ABSTRACT

A virtualization infrastructure that allows multiple guest partitions to run within a host hardware partition. The host system is divided into distinct logical or virtual partitions and special infrastructure partitions are implemented to control resource management and to control physical I/O device drivers that are, in turn, used by operating systems in other distinct logical or virtual guest partitions. Host hardware resource management runs as a tracking application in a resource management "ultravisor" partition, while host resource management decisions are performed in a higher level command partition based on policies maintained in a separate operations partition. The conventional hypervisor is reduced to a context switching and containment element (monitor) for the respective partitions, while the system resource management functionality is implemented in the ultravisor partition. The ultravisor partition maintains the master in-memory database of the hardware resource allocations and serves a command channel to accept transactional requests for assignment of resources to partitions. It also provides individual read-only views of individual partitions to the associated partition monitors. Host hardware I/O management is implemented in special redundant I/O partitions. Operating systems in other logical or virtual partitions communicate with the I/O partitions via memory channels established by the ultravisor partition. The guest operating systems in the respective logical or virtual partitions are modified to access monitors that implement a system call interface through which the ultravisor, I/O, and any other special infrastructure partitions may initiate communications with each other and with the respective guest partitions. The guest operating systems are modified so that they do not attempt to use the "broken" instructions in the x86 system that complete virtualization systems must resolve by inserting traps. System resources are separated into zones that are managed by a separate partition containing resource management policies that may be implemented across nodes to implement a virtual data center.